Sample Methods & Guidelines for Racial and Gender Pay Equity Department-Level Analysis

This document is a companion to "Racial and Gender Pay Gaps in California State Government: A Path Towards Workforce Equity," which provides an analysis of pay gaps across racial/ethnic and gender subgroups in California's civil service and a survey of strategies to advance parity within the State's workforce. Please reference the original report for additional discussion of context, analysis opportunities, and limitations.

Prepared for:
The California Health in All Policies (HiAP) Task Force

Prepared by:
Hinnaneh A. Qazi, MPP

July 2018

Table of Contents

Overview

In support of the Government Alliance on Race and Equity (GARE) Capitol Cohort, this document provides a sample of the type of statistical analysis that your GARE team can conduct to examine racial and gender pay inequities within your department. The overall goal of this analysis is to evaluate racial/ethnic and gender pay differences within your individual department using data available via the Management Information Retrieval System (MIRS). This document also provides tips and guidelines to ensure analyses and results are valid.

Data Retrieval

Gather employee microdata from your department's human resources' Management Information Report System (MIRS). The following variables should be pulled for each employee:

- Race/ethnicity
- Sex
- Pay (including base pay AND plus salary, if applicable)
- Classification
- Time base
- Tenure
- Pay period
- Age, DOB, or year of birth

Note: The purpose of this resource is to support departments to obtain an exploratory and high-level understanding of pay inequities. It is not an exhaustive list of all possible approaches or methods which can be leveraged to analyze racial/ethnic, gender, or pay trends within departments. We encourage departments to use this as a starting point, to further investigate

the scope and drivers of pay disparities, and to follow-up on their department's unique findings with additional research.

Data Preparation: Pre-Analysis

- The validity and efficacy of these methods hinges upon the overall department size and sample sizes of various racial/ethnic and gender subgroups.
  - In order to have confidence in your results, we recommend that you have at least 100 observations (e.g. employees) per category or subgroup. In some cases, it may be acceptable to combine subgroups (e.g., combining Asians with Pacific Islanders) to achieve more stable estimates.
  - For smaller departments, we encourage consulting research/data analysts or statisticians to ensure adequate sample sizes are available for some or all elements of this analysis. In addition, small sample sizes require extra considerations to ensure that confidentiality of employees is protected. A potential alternative to analyzing at a more granular-level is to simply analyze trends across White employees vs. employees of color.
- The following are recommendations to ensure that 1) there are "apples-to-apples" comparisons between subgroups and 2) findings are not inadvertently skewed:
  - Sum base pay and plus salary to estimate total monthly pay for employees (in other words, be sure to account for pay differentials).
  - Calculate pay gaps and average pay by race/ethnicity across all tenure and time base types *and* only among full-time permanent status employees.
    - Having both estimates will allow your department to determine whether certain racial/ethnic groups having a higher probability of being in limited term, temporary, or part-time roles contributes to pay disparities.
  - Standardize hourly or daily pay to be on a monthly pay period.
    - Not doing so will skew results (e.g., comparing a $17 per hour wage with a $4000 per month pay). One method of standardization is to identify all hourly or daily paid employees and multiplying their wage by the hours needed to work full time (example, [hourly wage] x 40 hours per week x 4 weeks per month). The alternative would be to gather data on hours and weeks worked per employee.
  - Drop major outliers. For example, most departments should exclude the department Director from analysis.

Sample Methods

Statistical analysis code in Stata (Do File) provided in addition.

STEP 1: CLEAN AND PREPARE DATA FOR ANALYSIS

Purpose: To prepare MIRS report/data set and create necessary variables for racial/ethnic and gender pay analysis.

*[1]Note: MIRS reports/excel sheets have many unnecessary spaces after observation entries, making it difficult for Stata to interpet values and generate variables. To address this, copy and paste values into excel and use "trim" function to remove spaces, then upload clean data set into Stata (or other statistical analysis software) for analysis.

*Clean up variable names
*Note: Original variable names from MIRS may vary by department.

```
rename ClassTitle class
rename Ethnic race
rename Gen sex
rename positionnumber positionno
rename salarytotal basepay
rename Salper payperiod
rename Tbase tbase
rename Ten tenure
```

*Drop observations with no recorded race/ethnicity.

```
drop if race=="W"
drop if race=="Z"
drop if race=="."
```

*Create dummy variables for race/ethnicity and sex.
*Note: URM stands for underrepresented minority (American Indians, Blacks, and Hispanic/Latinos). It is recommended that departments create other dummy variables that reflect unique department trends. For example, in some departments, Asian employees are highest paid. In this case, it may make sense to create a "POC without Asian" category to examine pay disparities between these groups as well.

---

[1] An asterisk (*) at the start of a line indicates text that is not Stata code, and appears before any text that explains steps and/or is meant to provide the reader more info before the subsequent code.

```
gen amerindian = race=="AMER INDIAN"
gen asian = race=="ASIAN" | race=="FILIPINO"
gen black = race=="BLACK"
gen hispanic = race=="HISPANIC"
gen othermulti = race=="OTHER"
gen pi = race=="PACIFIC ISL"
gen white = race=="WHITE"
gen poc = white==0
gen poc_urm = amerindian==1 | black==1 | hispanic==1

gen female = sex=="F"
gen male = sex=="M"
```

*Generate additional "other" dummy variable in the event sample sizes for some subgroups too small (Other department-level analyses have collapsed Pacific Islander, American Indian, and Other and Multiple race employees, but different collapsed subgroups may be better suited for other departments).

```
gen other = othermulti==1 | pi==1 | amerindian==1
```

*Generate dummy variables for age based on generational cohorts (can use other age thresholds depending on department distribution and/or interest in other trends)
*Note: these ranges are based on age in the year 2018.

```
gen geny = age<=21
gen millennial = age>=22 & age<=37
gen genx = age>=38 & age<=53
gen boomer = age>=54 & age<=72
gen silent = age>=73 & age<=90
gen over65 = age>=65
```

*Generate interaction terms for race/ethnicity, sex, and age.

```
gen amerindianfemale = amerindian*female
gen amerindianmale = amerindian*male

gen asianfemale = asian*female
gen asianmale = asian*male
```

```
gen blackfemale = black*female
gen blackmale = black*male

gen hispanicfemale = hispanic*female
gen hispanicmale = hispanic*male

gen othermultifemale = othermulti*female
gen othermultimale = othermulti*male

gen pifemale = pi*female
gen pimale = pi*male

gen whitefemale = white*female
gen whitemale = white*male

gen pocfemale = poc*female
gen pocmale = poc*male

gen poc_urmfemale = poc_urm*female
gen poc_urmmale = poc_urm*male

gen otherfemale = other*female
gen othermale = other*male

gen genywhite = white*geny
gen genypoc = poc*geny

gen millennialwhite = white*millennial
gen millennialpoc = poc*millennial

gen genxwhite = white*genx
gen genxpoc = poc*genx

gen boomerwhite = white*boomer
gen boomerpoc = poc*boomer

gen silentwhite = white*silent
```

```
gen silentpoc = poc*silent

gen over65white = white*over65
gen over65poc = poc*over65

*Create total monthly pay variable (base pay + plus salary)

gen totalpay = basepay+plussalary

*Examine skewness of pay to determine if should exclude outliers such as Department
or Agency director

histogram totalpay

*Create dummy variable for full-time permanent monthly salaried employees

gen fulltime = base=="FT" & tenure=="P"
tab fulltime
```

STEP 2: EXAMINE DEMOGRAPHIC BREAKDOWN OF DEPARTMENT

Purpose: To allow department to compare organization demographics to California civil service demographic breakdown, and California population demographic breakdown. This will also allow your department to understand how differences in demographics may also lead to differences in pay disparities as compared to the State civil service as a whole.

```
*Tabulate demographic distribution of department - to compare to California civil
service demographic breakdown, and California population demographic breakdown .

tab amerindian sex, column
tab asian sex, column
tab black sex, column
tab hispanic sex, column
tab othermulti sex, column
tab pi sex, column
tab white sex, column

histogram age
```

```
tab geny white, column
tab millennial white, column
tab genx white, column
tab boomer white, column
tab silent white, column
tab over65 white, column
```

STEP 3: ANALYZE PAY GAPS AND SKEWNESS, AVERAGE PAY OF RACIAL/ETHNIC AND GENDER GROUPS, & EXPLORE AGE EFFECTS

Purpose: To calculate average pay for all racial/ethnic and sex groups within the department, and subsequently estimate pay gaps between groups. Skewness of pay between white employees and employees of color will also be determined (i.e. are certain subgroups more likely to be concentrated in higher brackets of pay?). Another purpose of this step is to estimate the gender pay gap within the department as a point of comparison to gaps across state government as a whole (20.5%), allowing the department to have an idea of how inequities compare both along racial/ethnic and gender lines. This step will also allow your department to gauge how age correlates with pay, and if that explains any racial/ethnic pay disparities.

*Analyze gender pay gap as point of comparison.
*Note: Calculate among all employees and full-time permanent employees as well.

```
sum totalpay if female==1, detail
sum totalpay if male==1, detail

sum totalpay if female==1 & fulltime==1, detail
sum totalpay if male==1 & fulltime==1, detail
```

*Create histograms of total pay, to visualize White vs. people of color pay distribution and skewness.

```
twoway (histogram totalpay if poc==1, percent) ///
 (histogram totalpay if white==1, percent)


 twoway (histogram totalpay if poc==1 & fulltime==1, percent) ///
 (histogram totalpay if white==1 & fulltime==1, percent)
```

*Calculate average pay among all employees, by race/ethnicity and sex.

*HOW TO CALCULATE PAY GAPS: Take the average monthly pay of the lower paid group, divide that by the higher paid group's monthly pay, and multiply by 100 [(Lower paid group pay/Higher paid group pay)*100]=X. This will provide you the percent difference between each group. Example: Group A [lower paid group] is paid X% less than group B [higher paid group], on average.

*Note: Consider examining median pay across all employees and subgroups. Some pay gap analyses rely on median pay instead of average pay, however, median estimates will not provide confidence intervals or standard errors to distinguish whether certain groups are statistically different from one another.

mean totalpay
mean totalpay if fulltime==1
mean totalpay if tenure=="P" & tbase=="FT"
mean totalpay if tenure=="P" & tbase=="FT" & payperiod=="M"

*Median Pay Across All Employees
sum totalpay, detail
sum totalpay if fulltime==1, detail

*Average Pay by Subgroup, All Employees Included
mean totalpay if amerindian==1
mean totalpay if asian==1
mean totalpay if black==1
mean totalpay if hispanic==1
mean totalpay if othermulti==1
mean totalpay if pi==1
mean totalpay if white==1
mean totalpay if poc==1
mean totalpay if poc_urm==1

mean totalpay if amerindianfemale==1
mean totalpay if asianfemale==1
mean totalpay if blackfemale==1
mean totalpay if hispanicfemale==1
mean totalpay if othermultifemale==1
mean totalpay if pifemale==1
mean totalpay if whitefemale==1
mean totalpay if female==1

```
mean totalpay if pocfemale==1
mean totalpay if poc_urmfemale==1


mean totalpay if amerindianmale==1
mean totalpay if asianmale==1
mean totalpay if blackmale==1
mean totalpay if hispanicmale==1
mean totalpay if othermultimale==1
mean totalpay if pimale==1
mean totalpay if whitemale==1
mean totalpay if male==1
mean totalpay if pocmale==1
mean totalpay if poc_urmmale==1


*Average Pay by Subgroup, Restricted to Full-time Permanent Status Employees
mean totalpay if amerindian==1 & fulltime==1
mean totalpay if asian==1 & fulltime==1
mean totalpay if black==1 & fulltime==1
mean totalpay if hispanic==1 & fulltime==1
mean totalpay if othermulti==1 & fulltime==1
mean totalpay if pi==1 & fulltime==1
mean totalpay if white==1 & fulltime==1
mean totalpay if poc==1 & fulltime==1
mean totalpay if poc_urm==1 & fulltime==1


mean totalpay if amerindianfemale==1 & fulltime==1
mean totalpay if asianfemale==1 & fulltime==1
mean totalpay if blackfemale==1 & fulltime==1
mean totalpay if hispanicfemale==1 & fulltime==1
mean totalpay if othermultifemale==1 & fulltime==1
mean totalpay if pifemale==1 & fulltime==1
mean totalpay if whitefemale==1 & fulltime==1
mean totalpay if female==1 & fulltime==1
mean totalpay if pocfemale==1 & fulltime==1
mean totalpay if poc_urmfemale==1 & fulltime==1


mean totalpay if amerindianmale==1 & fulltime==1
mean totalpay if asianmale==1 & fulltime==1
```

```
mean totalpay if blackmale==1 & fulltime==1
mean totalpay if hispanicmale==1 & fulltime==1
mean totalpay if othermultimale==1 & fulltime==1
mean totalpay if pimale==1 & fulltime==1
mean totalpay if whitemale==1 & fulltime==1
mean totalpay if male==1 & fulltime==1
mean totalpay if pocmale==1 & fulltime==1
mean totalpay if poc_urmmale==1 & fulltime==1
```

*Examine effects of age on pay. Model below will reveal average pay for White employees and Employees of color by generational cohort.
*Note 1: May want to disaggregate at a more granular level (i.e. by race/ethnicity or POC identity + sex) if sample sizes within generational cohorts are large enough.
*Note 2: If no Gen Y'ers within department, omit "geny" terms from regression and begin with Millennial generation.
*Note 3: May also restrict this model to full-time permanent status employees if large enough sample sizes per age/race group.

```
reg totalpay pocgeny millennialwhite millennialpoc genxwhite genxpoc boomerwhite
boomerpoc silentwhite silentpoc
```

*Conduct statistical tests of significance between white employees and employees of color, and other groups as relevant (For example: Included below is a test of significance between White employees vs. employees of color and Other and Multiple race employees, inclusive of all employees and then restricted to full-time permanent employees.).

```
ttest totalpay, by(white)
ttest totalpay if white==0, by(othermulti)
```

```
ttest totalpay if fulltime==1, by(white)
ttest totalpay if white==0 & fulltime==1, by(othermulti)
```

STEP 4: EVALUATE SOURCES OF INEQUITIES: TENURE TYPE, TIME BASE, ACROSS-CLASSIFICATION DIFFERENCES, AND POTENTIAL WITHIN-CLASSIFICATION PAY DISPARITIES

Purpose: To analyze potential drivers of pay inequities within the department, and determine whether certain racial/ethnic and sex subgroups are disproportionately and systematically overrepresented in position types which may lead to disparities. These variables (tenure type,

time base, and classification) were selected due to their availability in MIRS, but many other variables (such as occupation or supervisor status) can be evaluated as well.

*Sources of inequities - TENURE TYPE.

*Tabulate how many employees exist per tenure type
tab tenure

*Approximate mean total pay by tenure type to assess whether certain subgroups concentrated in lower vs. higher paid tenure types.

mean totalpay if tenure=="C"
mean totalpay if tenure=="L"
mean totalpay if tenure=="P"
mean totalpay if tenure=="R"
mean totalpay if tenure=="T"

*Tabulate racial/ethnic and sex breakdown by tenure type (Career Executive Assignment = C, Limited Term = L, Permanent = P, Retired Annuitant = R, Temporary = T).

tab race tenure, column
tab sex tenure, column

*Examine demographic breakdown of dept-level leadership (Career Executive Assignments).

tab race sex if tenure=="C"

*Sources of inequities - TIME BASE.

*Tabulate how many employees exist per time base
tab tbase
tab tbase if tenure=="P"

*Approximate mean total pay by time base to gauge whether certain subgroups concentrated in full-time vs. intermittent/temporary work.

*Note: Department may want to isolate and examine trends across other time bases [e.g. 001/002] if enough employees.

mean totalpay if tbase=="FT"
mean totalpay if tbase=="INT"
mean totalpay if tbase!="FT" & tbase!="INT"

*Tabulate racial/ethnic and sex breakdown by tenure type.

tab race tbase, column
tab sex tbase, column

*Evaluate whether white employees employed in higher paying intermittent/temporary work.

mean totalpay if tbase!="FT" & white==1
mean totalpay if tbase!="FT" & white==0

*Sources of inequities - ACROSS CLASSIFICATION DIFFERENCES.

*Generate pay quintiles (lowest 20% paid, 20 to 40% lowest paid...top 20% paid) of classifications to determine distribution of racial/ethnic groups across class quintiles.
*NOTE: Follow steps 1-6 to determine the quintile "rank" of each classification:

*1. Use collapse function in Stata [collapse totalpay, by(class)] (WARNING: Do not override original data set by saving new collapsed data set over original data set)
*2. Export Stata data set into excel file
*3. Use percentile [=PERCENTILE.INC(A1:AX,0.2)] function in excel to determine quintile thresholds (20%, 40%, etc.) of classification pay
*4. Sort classes into quintile groups, and create column to list rank (i.e. 1 [lowest paid quintile], 2 [second lowest paid quintile], etc.)
*5. Merge [merge m:1 class using "disk hyperlink"] quintile data (i.e. which quintile each class falls into) from excel with original data set in stata
*6. Rename new variable which has quintile ranking (this analysis renamed the variable "classquintile")

*Tabulate the sex and racial/ethnic breakdown of each subgroup per class quintile.

```
tab female classquintile, column
tab male classquintile, column


tab asianfemale classquintile, column
tab asianmale classquintile, column
tab asian classquintile, column


tab blackfemale classquintile, column
tab blackmale classquintile, column
tab black classquintile, column


tab hispanicfemale classquintile, column
tab hispanicmale classquintile, column
tab hispanic classquintile, column


tab otherfemale classquintile, column
tab othermale classquintile, column
tab other classquintile, column


tab whitefemale classquintile, column
tab whitemale classquintile, column
tab white classquintile, column
```

*Sources of inequities - POTENTIAL WITHIN CLASSIFICATION DIFFERENCES.

*Determine largest classes within department.

```
tab class, sort
```

*Tabulate racial/ethnic and sex demographic breakdown of classification. Be wary of interpreting results for small subgroups within-classification.

```
tab race sex if class=="[CLASSIFICATION TITLE]"
```

*Estimate mean pay for all employees within classification, and for subgroups. It is strongly urged to look at average pay within the context of pay ranges for the classification.

mean totalpay if class=="[CLASSIFICATION TITLE]"

mean totalpay if class=="[CLASSIFICATION TITLE]" & white==1
mean totalpay if class=="[CLASSIFICATION TITLE]" & white==1 & female==1
mean totalpay if class=="[CLASSIFICATION TITLE]" & white==1 & male==1

mean totalpay if class=="[CLASSIFICATION TITLE]" & asian==1
mean totalpay if class=="[CLASSIFICATION TITLE]" & asian==1 & female==1
mean totalpay if class=="[CLASSIFICATION TITLE]" & asian==1 & male==1

mean totalpay if class=="[CLASSIFICATION TITLE]" & black==1
mean totalpay if class=="[CLASSIFICATION TITLE]" & black==1 & female==1
mean totalpay if class=="[CLASSIFICATION TITLE]" & black==1 & male==1

mean totalpay if class=="[CLASSIFICATION TITLE]" & hispanic==1
mean totalpay if class=="[CLASSIFICATION TITLE]" & hispanic==1 & female==1
mean totalpay if class=="[CLASSIFICATION TITLE]" & hispanic==1 & male==1

mean totalpay if class=="[CLASSIFICATION TITLE]" & other==1
mean totalpay if class=="[CLASSIFICATION TITLE]" & other==1 & female==1
mean totalpay if class=="[CLASSIFICATION TITLE]" & other==1 & male==1

mean totalpay if class=="[CLASSIFICATION TITLE]" & female==1
mean totalpay if class=="[CLASSIFICATION TITLE]" & male==1

mean totalpay if class=="[CLASSIFICATION TITLE]L" & poc==1
mean totalpay if class=="[CLASSIFICATION TITLE]" & poc==1 & female==1
mean totalpay if class=="[CLASSIFICATION TITLE]" & poc==1 & male==1

mean totalpay if class=="[CLASSIFICATION TITLE]" & poc_urm==1
mean totalpay if class=="[CLASSIFICATION TITLE]" & poc_urm==1 & female==1
mean totalpay if class=="[CLASSIFICATION TITLE]" & poc_urm==1 & male==1

*Other potential interesting within-classification analyses:

*Control for service months within a class or how long an employee has been within the department.
*Examine pay by fulltime vs. non-fulltime employees within class.

*Analyze how average employees of color pay changes after omitting higher paid minorities (such as Other and Multiple race employees in this example).
*Explore other intersections between race/ethnicity and sex (e.g. how are females of color paid in comparison to males of color?).

```
mean totalpay if class=="[CLASSIFICATION TITLE]" & white==1 & fulltime==1
mean totalpay if class=="[CLASSIFICATION TITLE]" & white==0 & othermulti==0 &
fulltime==1

mean totalpay if class=="[CLASSIFICATION TITLE]" & white==1
mean totalpay if class=="[CLASSIFICATION TITLE]" & white==0
mean totalpay if class=="[CLASSIFICATION TITLE]" & white==0 & othermulti==0

mean totalpay if class=="[CLASSIFICATION TITLE]" & pocfemale==1
mean totalpay if class=="[CLASSIFICATION TITLE]" & pocmale==1
mean totalpay if class=="[CLASSIFICATION TITLE]" & whitefemale==1
mean totalpay if class=="[CLASSIFICATION TITLE]" & whitemale==1

ttest totalpay if class=="[CLASSIFICATION TITLE]", by(white)
ttest totalpay if class=="[CLASSIFICATION TITLE]", by(other)
ttest totalpay if class=="[CLASSIFICATION TITLE]" & poc_urm==1, by(sex)
```

*Repeat above code for additional classifications if large enough class sizes available.

Future Research Questions

Future research should also consider investigating:

1. If large enough sample sizes are available:
   a. Additional classifications to confirm whether there are within-classification pay differences.
   b. Estimate pay within-classification while controlling for service months to determine whether there are some inequities in terms of merit-based salary adjustments.
2. Racial/ethnic and sex distribution across occupations and supervisor vs. rank and file roles.
   a. Occupation group information per classification can be retrieved from CalHR 5102 Report – State Employees by Occupational Group and Classification.

b. Supervisor/managerial classification can be pulled from [publicly available pay scale range](#) information using CBID codes.

3. Evaluate additional intersections between race/ethnicity, sex, and age. For example, what are pay gaps between younger women of color versus older white men in the department?

Supplemental Materials

In addition to this memo, the following are attached:

*   A Stata Do File for statistical analyses. This should be easily replicable in other statistical analysis platforms such as SPSS, R, etc.
*   A data glossary for variables and terms in the statistical analysis code.

Questions?

For questions please contact:

Meredith Lee, Office of Health Equity, California Department of Public Health, Health in All Policies Team Lead at [Meredith.Lee@cdph.ca.gov.](mailto:Meredith.Lee@cdph.ca.gov)